

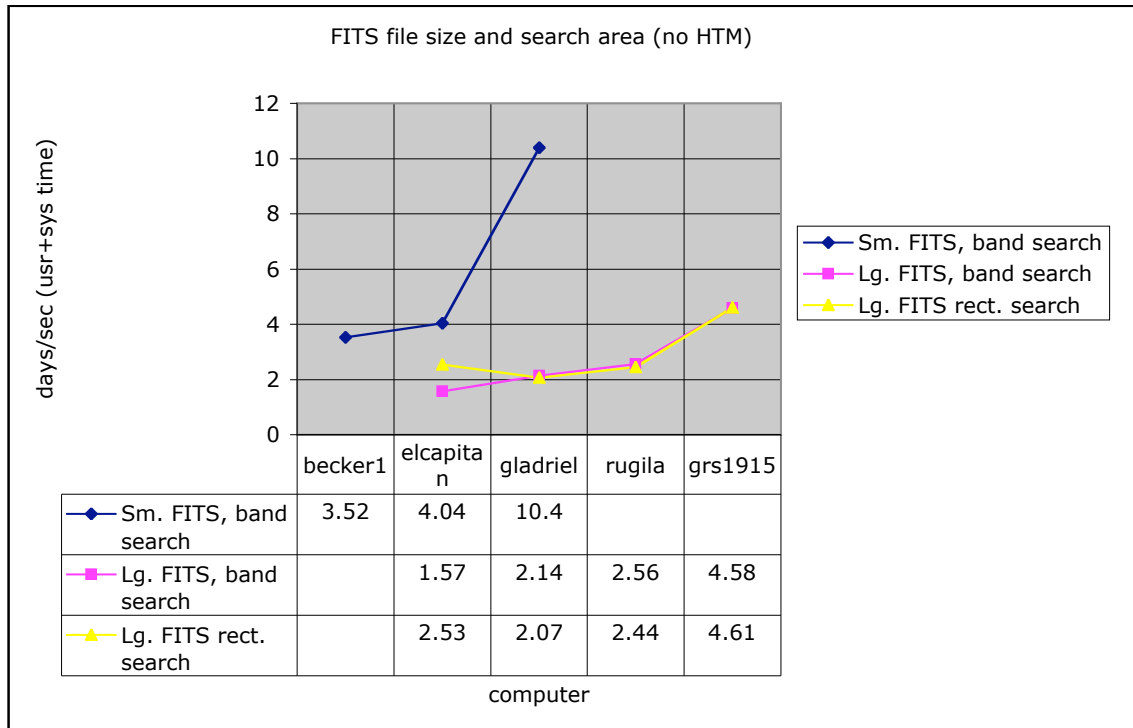
FITS File Data Searching Benchmark Tests

Introduction.

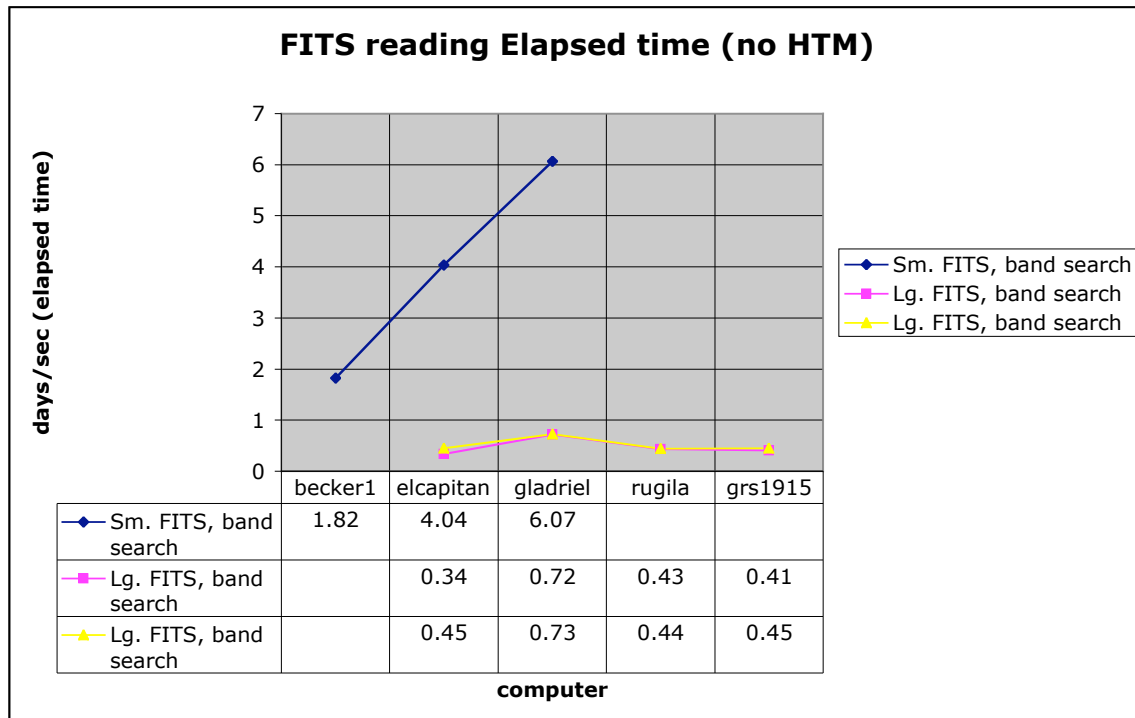
Since Pat Nolan found that searching FITS files was faster than querying freeware databases (MySQL and PostgreSQL), we decided to explore the speed of searching FITS files. This was done in a effort to get a more accurate picture of the database speed for hardware software design.

Results:

1. Search speed in days per second, where seconds is the number of user cpu and related system time. This compares speeds of searching mini-FITS files made by Seth with the Seth's files bloated out to their full size. It also shows the effect of the size of the written search results on performance.

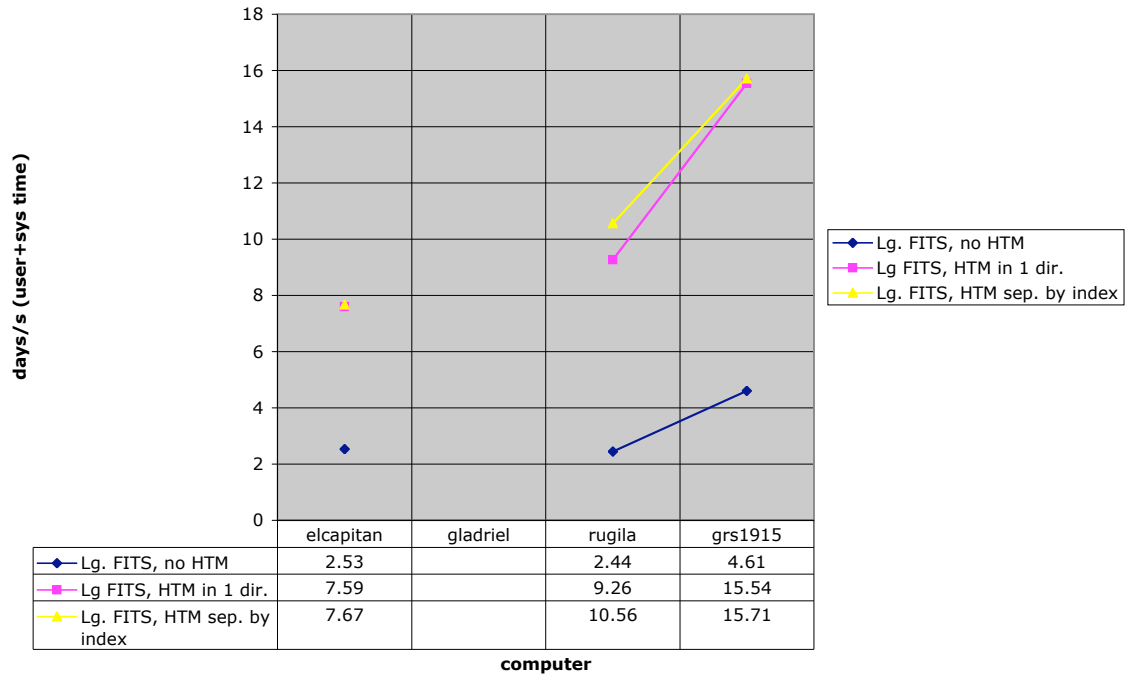


2. Search speed in days per second, where seconds is the total elapsed time. This compares speeds of searching mini-FITS files made by Seth with the Seth's files bloated out to their full size. It also shows the effect of the size of the written search results on performance.

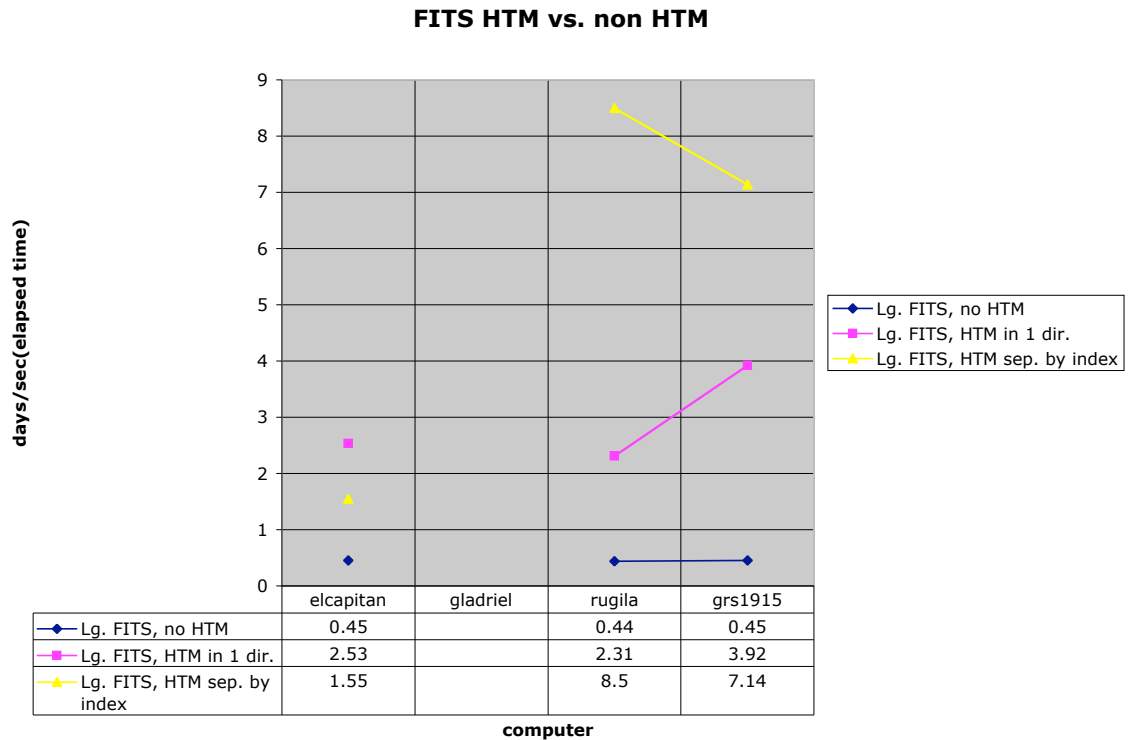


3. Here we consider the effect of breaking up the data into 8 HTM pixels. Since now the files are 1/8 the size, does search speed increase by 8? (No – more like 4). What is the effect of getting the files separated by index and put sequentially on the disk – is that significantly faster than just keeping them all in one directory? (doesn't seem to make too much difference. Again the Y axis is days/sec (user CPU + related system time)

FITS HTM vs. not HTM



4. Same as 3) but using elapsed time. Here we consider the effect of breaking up the data into 8 HTM pixels. Since now the files are 1/8 the size, does search speed increase by 8? (No – more like 4). What is the effect of getting the files separated by index and put sequentially on the disk – is that significantly faster than just keeping them all in one directory? (doesn't seem to make too much difference. The Y axis is days/sec of elapsed time.



Computers used:

1) becker1 - the head node of the Beowulf used for prototype performance studies. The Beowulf is made of Intel Celeron 350 and 400 MHz processors with 17 Gb IDE drives (could be 5400 rpm drives) - not sure.

2) elcapitan - a Pentium III linux box with IDE disks processor clock speed not known.

3) gladriel - a AMD Athlon 1900+ dual processor 1.6 GHz with 7200 rpm IDE disk drives

5) rugila - a near clone of gladriel processors and disks are the same, some peripherals are different - sound card network card...

5) grs1915 - a RH Linux machine with SCSI disks.